EXPLORING SPATIAL OPTIMIZATION TECHNIQUES FOR THE PLACEMENT OF FLOW MONITORS UTILIZED IN RDII STUDIES

Christopher A. Skehan

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Geography,
Indiana University

July 2010

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science

Aniruddha Banerjee, Ph.D., Chair

Jeffery S. Wilson, Ph.D.

Master's Thesis Committee

Daniel P. Johnson, Ph.D.

Acknowledgements

I am heartily thankful to my thesis committee, Aniruddha Banerjee, Ph.D, Jeffrey S. Wilson, Ph.D, and Daniel P. Johnson, Ph.D whose guidance and support from the initial to the final level enabled me to develop an advanced understanding of the subject.

I would also like to include my sincerest thanks to Patrick Stevens of ADS LLC, who offered professional support during the course of this study and provided a number of documentation needed to complete this study.

Lastly, to my wife Jill, for seeing my true potential to accomplish such a great task. I could fill an entire book with the encouragement, patience, and love you have provided throughout this process.

Table of Contents

Introduction

In 1876, Boston, MA constructed one of the first combined sewer collection systems. This system was highly recognized and credited as one of the finest wastewater and stormwater collection systems within the United States. By 1980, a disregard and failure to conduct regular maintenance caused a near system-wide collapse of the sewer collection system (MWRA Online, 2009). As failures increased across the country, the United States Environmental Protection Agency (USEPA) increased the regulation, management, and maintenance of sanitary wastewater collection systems (USEPA, 2004). These changes, which have occurred primarily within the last 30 years, have pressured and forced municipalities to execute improvements and repairs, especially in older collection systems.

The aging infrastructure of a wastewater collection system can leak, capture ground water, and capture precipitation runoff. These are some of the most common problems in many of today's US collection systems and are often collectively referred to as Rain Derived Inflow and Infiltration (RDII or I/I). RDII can cause a number of significant problems for a local municipality (SSOAP Toolbox, 2009). According to the USEPA, "RDII is the main cause of sanitary sewer overflows (SSOs) to basements, streets, or nearby receiving waters and can also cause serious operating problems at wastewater treatment facilities" (SSOAP Toolbox, 2009). Hence, there is both national and local interest in developing strategies for measuring, reducing, and eliminating RDII in municipal collection systems.

Municipalities often engage in a number of methods to tackle RDII, including smoke testing, sewer televising, and flow monitoring. A recent study has suggested that flow monitoring is one of the most common and cost effective approaches for measuring and studying RDII (Stevens, 2005). The primary goal of flow monitoring is to record wastewater data, such as the depth, velocity, and quantity within a given section of the sewer collection system over a unit of time. Engineers frequently compare the data captured by a flow monitoring device with rainfall or storm event data. Inflow and infiltration patterns within a collection system can then be extracted from these results. A detailed manuscript by Mitchell and Stevens (2005) provides a list of additional considerations for utilizing a flow monitoring study.

Before any data can be collected from these flow monitoring devices, it is necessary to determine how many monitors to utilize in the study and where to best place these monitors within the wastewater collection system. There are some general guidelines available to assist with choosing the appropriate number of monitors to utilize in an RDII study. Stevens (2005) suggests that, for optimal results, a flow monitor should capture data for a unit area which contains 10,000 to 15,000 linear feet (LF) of sewer pipe. However, there is a limited amount of literature that provides the framework for determining the placement of a flow monitor.

It is also necessary to introduce the full complexity of this process and how a flow monitor's placement is currently determined. One must first consider the conundrum that exists between the number of monitors and their placement within a large network. The number of flow monitors available for a study can directly influence the placement of the monitors. In addition, the known placement of a flow monitor within the system can also influence the number of flow monitors utilized in the study. The process which exists between the number of monitors and their placement is currently being solved using a non-optimized methodology.

Let us first consider how the number of flow monitors can influence the placement. Figure 1 illustrates a section of a sewer system, which is represented by nodes and vertexes. Each node represents a potential manhole access point where a flow monitor can be placed within the collection system network. Each vertex is a section of sewer pipe within the system. The symbols that represent each vertex have been adjusted to show the direction of flow within the network. In Figure 1, approximately 80 potential flow monitoring locations are available. Also in this example, there is approximately 60,000 linear feet of pipe, which is represented by the vertices. If we utilize the Stephens (2005) methodology, where each monitor will service 10K to 15k feet of pipe, then approximately 4 to 6 flow monitors would be needed to obtain optimal RDII results. Even with such a small example, where is the optimal placement of each monitor? How will the optimal placement change given 4 flow monitors versus 6 flow monitors? Unfortunately, these questions cannot simply be answered by utilizing the current methodology of Stevens (2005). This question becomes even more complex as the study area size increases.

Next, let us consider how the location of a flow monitor can lead to the influence of additional flow monitors. Figure 2 shows the same section of the sewer system, but this time, we have added the potential location of a flow monitor. Hypothetically, this location could be selected because of a public interest to eliminate a periodical sewer overflow that occurs with the arrival of a large rain event. The red star icon represents the potential location of the flow monitor, and the green shaded section represents the area of influence of pipe that will affect this monitor. This flow monitor has 19,000 linear feet of influence which, according to Stevens (2005), might not yield optimal results. The engineer then considers that the flow monitor's basin should be sub-divided into two smaller basins. The result will be something similar to Figure 3. In Figure 3, an additional flow monitor was added to optimize the RDII results and will assist in determining which subsection of sewer is causing the SSO. This concept is further explained in Stevens (2005), and similar examples are presented.

Non-Optimized Methods: Arbitrary, Local, and Sequential

It can be a complicated task determining a flow monitor's placement when the size and complexity of a sewage collection system increases. In a large municipal or urban collection system, hundreds of square miles of area and several million LF of pipe can be present. Therefore, it is a common a trend to adapt an arbitrary process in order to determine a flow monitors placement. As discussed in Stevens (2005), once the number of meters has been established, it is common practice for an individual to visually analyze and choose a location within the system to place a flow monitoring device. This process can be considered arbitrary, as it is highly unlikely that two persons will establish the same result for a monitor's placement and moreover for multiple monitor placements in larger systems.

Stevens (2005) methods also utilize more of a local optimization approach, or the method of choosing individual locations in a specific and local area. For example, an engineer will browse a collection system network and choose a potential flow monitoring location, considering only a single location at a time. This process usually relies on a map where a local and zoomed-in area is studied to determine a monitor's placement. This approach is currently be used today, however, it undermines a mathematical optimization approach which can be utilized within a GIS application.

Furthermore, Stevens (2005) technique is sequential, as a single result is determined one after another until the number of monitors utilized in the study is exhausted. This technique is time consuming, as each location must be determined individually. The overall size of the collection system can lead to additional complications, as thousands of different locations must be considered. Such techniques rarely lead to optimal flow monitor placement within the collection system.

Optimized Methods: Deterministic, Global, and Simultaneous

In recent years it has become common practice for larger municipalities to "digitize" their assets. Access to digital records, CAD files, or GIS data is necessary for a number of organizations to conduct their daily regular business. When a digital copy of a sewer collection system is available, and when the digital data has preserved the direction of flow of the sewer network, the principles of operations research and graph theory can be applied to solve a problem such as the placement of flow monitors for an RDII study within a given network.

The goal of operations research (OR) is to use mathematical and scientific approaches to solve problems often with the goal of finding an optimal result (Churchman et al, 1957). Kirby (2003) outlines how historically OR was utilized for transportation, communication, and strategy and planning for military operations and war. Today, the term loosely refers to principles that utilize a more peaceful approach for solving such problems. Operations research can be utilized for solving a facility location problem or a network partitioning problem. As we will later see, a facility location model coupled with network partitioning can help optimize a flow monitor's placement.

Within OR, graph theory and network flows can assist in solving this problem. Graph theory has been around for over 250 years, as Leonhard Euler first introduced this concept while trying to solve The Seven Bridges of Königsberg problem (Biggs et al, 1986). Additionally, network flows are commonly associated with graph problems, as the flow of a network is critical to the interactions between nodes and vertices within the graph. Many of these approaches are already available for users who conduct transportation and logistical studies. When a large collection system network is available, logistical approaches and methods can be adapted to suit this dilemma such as the flow monitor placement problem.

Operations Research and graph theory provide the tools necessary to solve this problem by utilizing a deterministic process, a global optimization approach, and a simultaneous technique. A deterministic process would be favorable as it would yield the same result every time, regardless of who is performing the study. Rather than solving for a single location and the placement within a localized area, we can utilize a global optimization approach that would analyze the entire system as a whole. Lastly, utilize a simultaneous technique, especially when a global optimization approach is utilized, would be favorable as all possible solutions would be considered and an optimal result could be produced.

The goal of this study is to investigate such optimized methods and their potential to improve flow monitor placement, especially for RDII studies, and to improve upon

Stevens (2005) methodology. This project will adopt a methodology from the "facility location problem", a branch of operations research and graph theory. Solutions to a facility location problem will be adapted and utilized within a transportation GIS application to determine optimal placement. While utilizing a facility location tool, this study will explore distance as the primary cost to optimize flow monitor placement. To compare non-optimized locations and optimized locations, network partitioning was used.

Design

This study has been designed to adapt and utilize a mathematical approach for solving a facility location problem. This study will adopt the facility location model which is available in the TransCAD routing and logistics software. TransCAD will also provide a model for applying a network partition to each location within the network. The facility location model used in this study uses the methodology of the *P-Median* problem outlined in Densham and Rushton (1992). The hardware utilized for this study was a PC with following specifications: 3.00 GHz Intel: Core 2 Dual Processor, and 3.25 GB of System Memory. These specifications have been provided as basis of comparison for future studies as the time associated with running the facility location model and network partitioning might change with different hardware

Data

Shapefiles of Pre-Existing Sewer Features: Lawrence, Indiana (2009)

A vector line file was used which contains 1,202,415 million linear feet of sanitary and combined collection system lines for Lawrence, Indiana. The file itself was not topologically integrated for the facility location problem, and as a result, the file was converted into a standard geographical file by the TransCAD software. This file format was used to create a network, which is necessary to solve this type of problem. There are approximately 25,000 nodes associated with this vector line file. As mentioned earlier, each node represents a potential location were a flow monitor can be placed.

Physical constraints within the built environment often cause access limitations when installing flow monitors. Therefore, it is common practice to use an additional node layer showing sewer manhole locations, which are the physical locations to gain access into a section of sewer collection system. Manhole locations are critical when determining the installation of a flow monitor. However, this study will not utilize the manhole layer, as one can easily perform a simple *closest location analysis* between the node and manhole layer within most GIS applications.

Non-Optimized Flow Monitoring Locations: Lawrence, Indiana (2009)

During a 2009 ADS RDII Study, 53 locations were hand selected by an engineer. These pre-determined locations were selected by applying a non-optimized methodology. As a result, arbitrary locations were chosen by an engineer, as each location was determined by applying a local optimization approach and a sequential placement technique. This data will be analyzed and compared to the results of this study by performing a network partition on each location and comparing the overall network influence that each location has within the network. Network partitioning can provide the results for all upstream LF that each flow monitoring location will measure as well as the total LF that will be measured within the entire network by all of the flow monitoring devices.

Methods

Determining the placement within a collection system is closely related to the facility location problem. Daskin (2008) has outlined this problem and its complexity and discusses several techniques that have been explored in recent years. The basic principle of the problem is usually the same, where the goal is often to explore an optimization between facilities and clients within a network. This optimization is simply an extension of the *P-Median* problem discussed in Christofides (1975). To solve this problem, there are several key characteristics that must first be determined. These include defining the network between the facilities and clients, calculating the connection costs, and calculating the facility costs. (Hajiaghayi et al, 2003).

A facility location model was setup to solve this problem. The GIS data shows a one-way network, where each pipe segment can only flow in one direction. Once the network was defined, it was then necessary to determine the connection costs between pipe segments, which were defined by the length between each segment within the network. Each pipe segment was also assigned a cost, which in this study was also reflected by length between nodes. In this study, the facility costs were also defined by the length between segments.

TransCAD has the ability to create as many “facilities” as needed to accommodate the entire network or gives the user an option to manually choose the number of facilities to place within the network. In both cases, the optimal placement of the monitor is obtained by adopting a mathematical optimization algorithm that utilizes the connections between each node, measured length between nodes, the network flow and the defined costs of each potential location. For this study, we manually chose 53 flow monitoring locations (facilities) to be determined using the facility location model. All 4,091 nodes were used as potential clients in the facility location model.

Once the flow monitoring locations were established by the facility location model, a network partitioning method was utilized on the original 53 non-optimized locations and the new 53 locations selected by the facility location model. The network partitioning showed which areas within the network would influence a particular location. This proved to be very useful for comparing results between the non-optimized locations and optimized locations.

Results

The following results compare two sets of data; the non-optimized locations, which were arbitrary selected by an individual, and the optimized locations, which were determined using a facility location model. The numbers discussed below are a result of the network partition which was performed on each set of locations. Table 1 shows the networking partitioning results for the non-optimized flow monitors versus the optimized flow monitors. Using the Stevens (2005) approach, the 53 non-optimized locations would monitor 519,218 LF of pipe within the network. Approximately 43% of the entire collection system is being monitored by these 53 hand-selected, non-optmized locations. Using the facility location model, the 53 optimized flow monitoring locations would monitor 755,817 LF of pipe within the network. Approximately 63% of the entire collection system could have been monitored if these locations were utilized during the RDII study.

Figure 5 and Figure 6 show the networking partitioning maps which were calculated for each flow monitor. Note that Figure 5 shows these network partitions for the non-optimized locations where as Figure 6 shows the network partitions for the optimized locations. Figure 7 shows the “island” areas, which are the disconnected pipe segments that were created as a result of converting the shapefile into a network file.

As an unintended result, the networking partitioning tool also provided a time saving feature for identifying the network basin influence for each flow monitor. Network partitions were utilized for extracting pipe length totals for each flow monitor. Using the Lawrence, Indiana sewer network and 53 locations selected, the networking partitioning results were extracted in less than 20 minutes for both the non-optimized and optimized locations.

Conclusion

This study explored a facility location model where 53 new flow monitoring locations were determined and then compared to 53 arbitrarily chosen locations. To evaluate changes between the non-optimized locations to the newly determined optimized locations, the network partitioning results were compared. The goal of this study was to show how OR and GIS could be utilized to optimize flow monitoring locations for RDII studies. Results show that the facility location model would have increased the total percentage of collection system which would be monitored by almost 20%. While the average upstream network influence increased to 14,000 LF for the optimized locations, we can see that this is still within the limits of Stevens (2005) methodology. In this study, spatial optimization was achieved by using only 53 locations. Also, as mentioned in the results section, the time needed to create and extract a flow monitor's LF and other basin information was greatly reduced. The approach outlined in this study uses a deterministic process, a global optimization approach, and a simultaneous technique to achieve spatial optimization for flow monitor placement within a collection system network.

The following considerations should be explored in future studies. First, the original digital pipe data had numerous connection errors. When there are connection errors in the data, the network is incomplete and can cause a number of analysis problems. The facility location model treated these disconnected sections as "islands", which affected the optimal placement for each flow monitor. (Figure 7) Connection improvements within the digital data might provide a higher accuracy of optimization.

Second, this study did not account for the influence between lift stations and the collection systems network. Lift stations manually direct and pump wastewater and stormwater to other locations within the network. The pipes that connect a lift station may not have been reflected in the data used in this study, as there was a limited amount of information provided with this digital data. A pump station layer could be merged with the pipe layer to provide a more complete wastewater or stormwater collection system network.

Third, when using Stevens (2005) methodology, multiple results should be considered. It would be beneficial to determine the upper and lower limits for the number of optimal flow monitors and compare how these would change using the facility location model.

Lastly, a pipe's diameter could be used to compute the connection cost associated within the facility location problem. This would provide an alternative method for defining the facility costs which are necessary to execute a facility location model.

	Number of Locations	Utilized Line Segments	Relative Efficiency	Total LF Monitored	% of System Monitored	Average LF for Each Flow Monitor
Non Optimized Locations	53	2067	66%	519,218	~43%	10,383
Optimized Locations	53	3136	100%	755,817	~63%	14,153

Table 1: Summary of Network Partitioning. The yellow highlighted areas were critical for comparing the spatial optimization techniques utilized in this study.

Figure 1

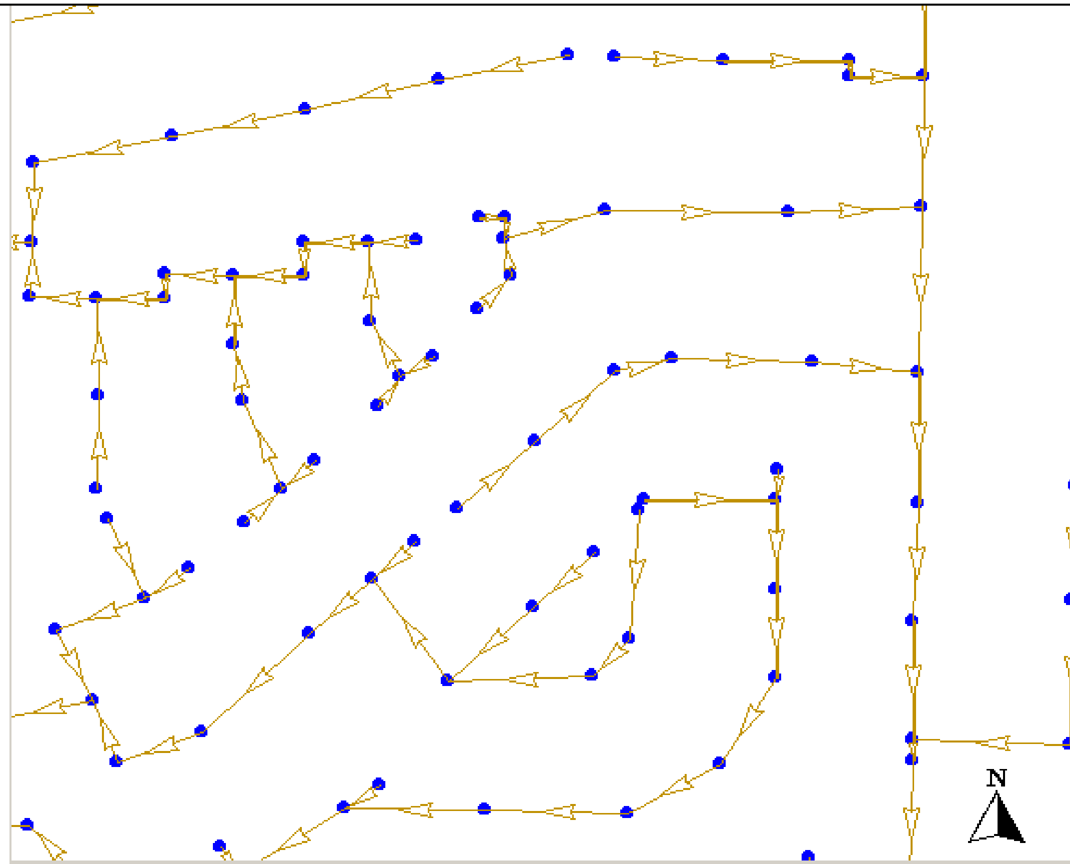


Figure 1 – A Section of Sewer Pipe in Indianapolis, IN – Each blue dot represents a node within the system, and the brown, directional lines represent the vertices. Notice that the direction of flow is represented by the vertices.

Figure 2

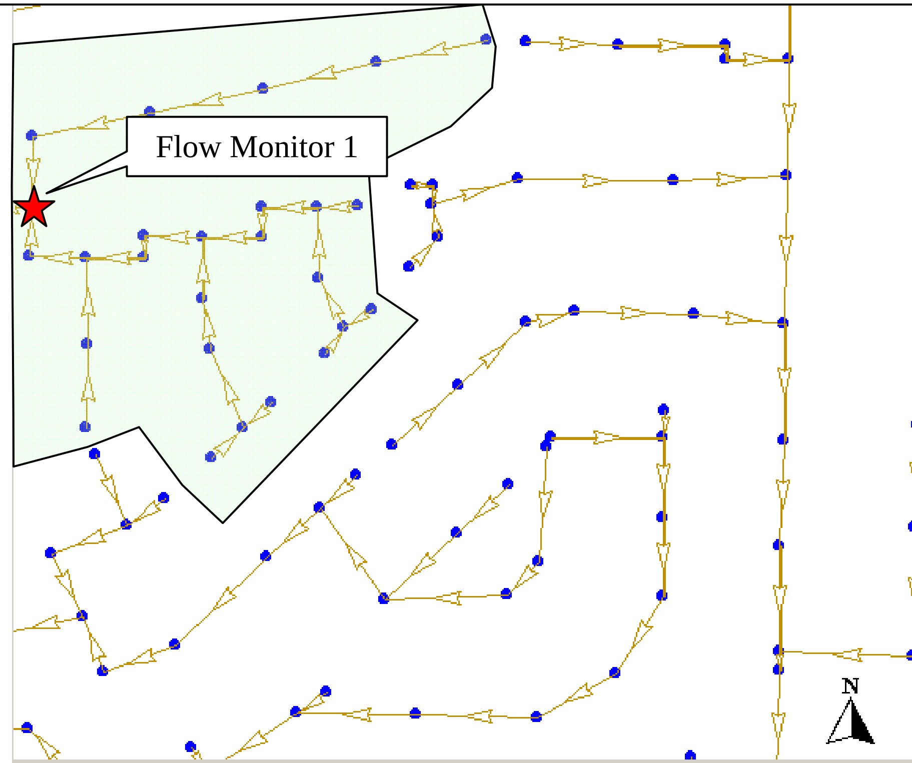


Figure 2 – The same section of sewer pipe in Indianapolis, IN – However, in this example we have highlighted with the star icon the potential location for a flow monitor's placement within the system for the study. The area of influence has also been highlighted in this example to show which sections of the network will influence the flow monitor.

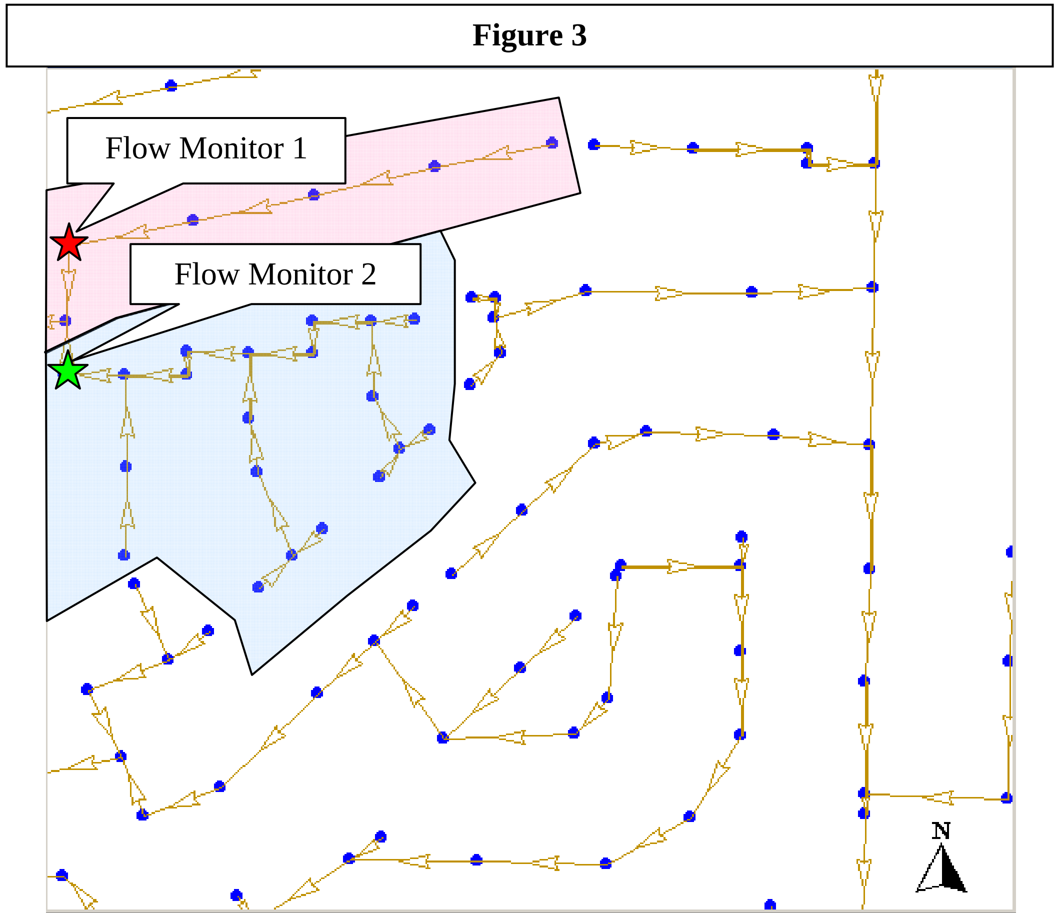


Figure 3 – After the original location was examined, it was determined that 2 flow monitors would be needed to obtain optimal results, where each flow monitor would be influenced by a sub-section of the original basin.

Figure 4

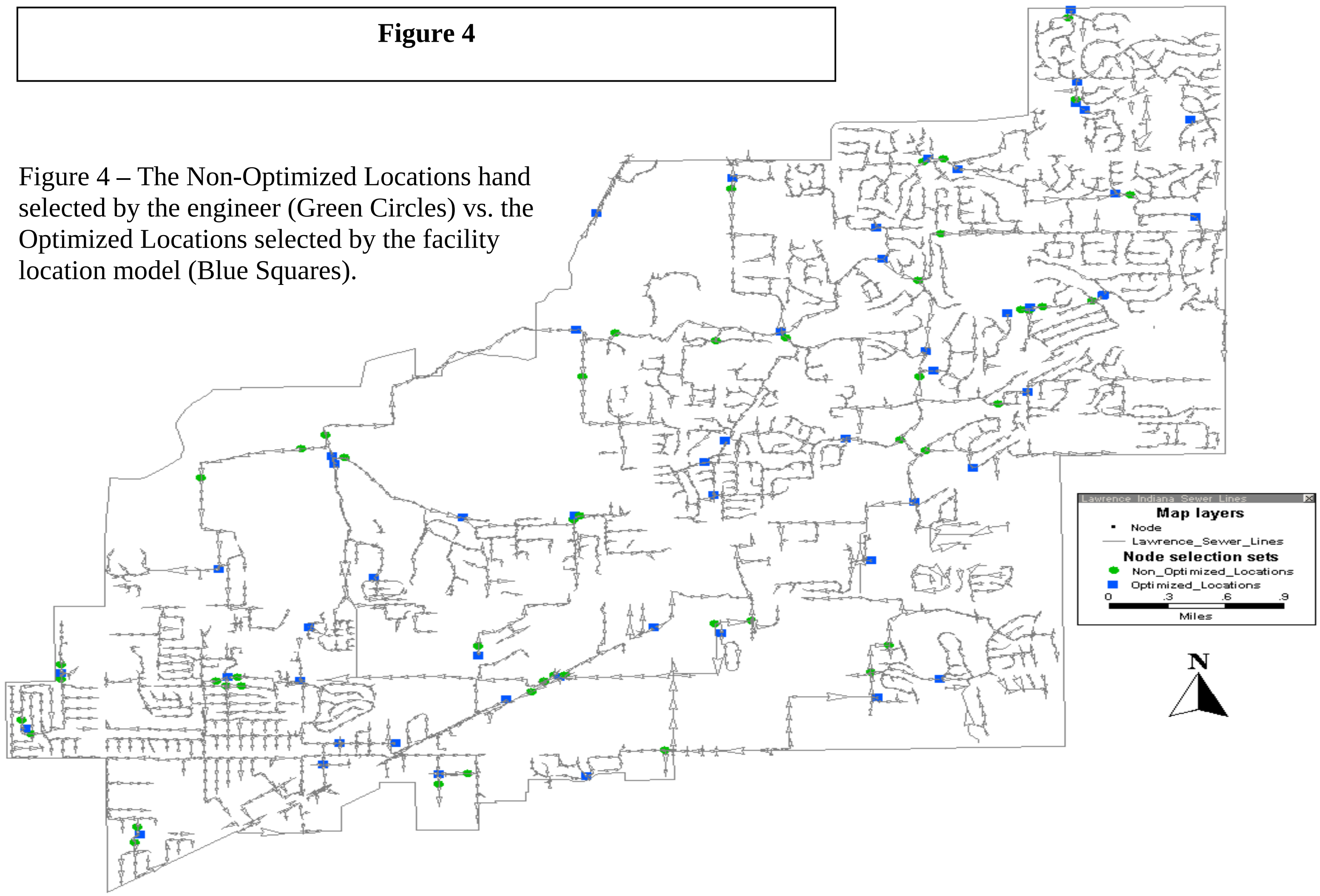


Figure 4 – The Non-Optimized Locations hand selected by the engineer (Green Circles) vs. the Optimized Locations selected by the facility location model (Blue Squares).

Figure 5

Figure 5 – The Non-Optimized Locations and the calculated network partition for each location. The boundary is also highlighted in black to show the area of study used.

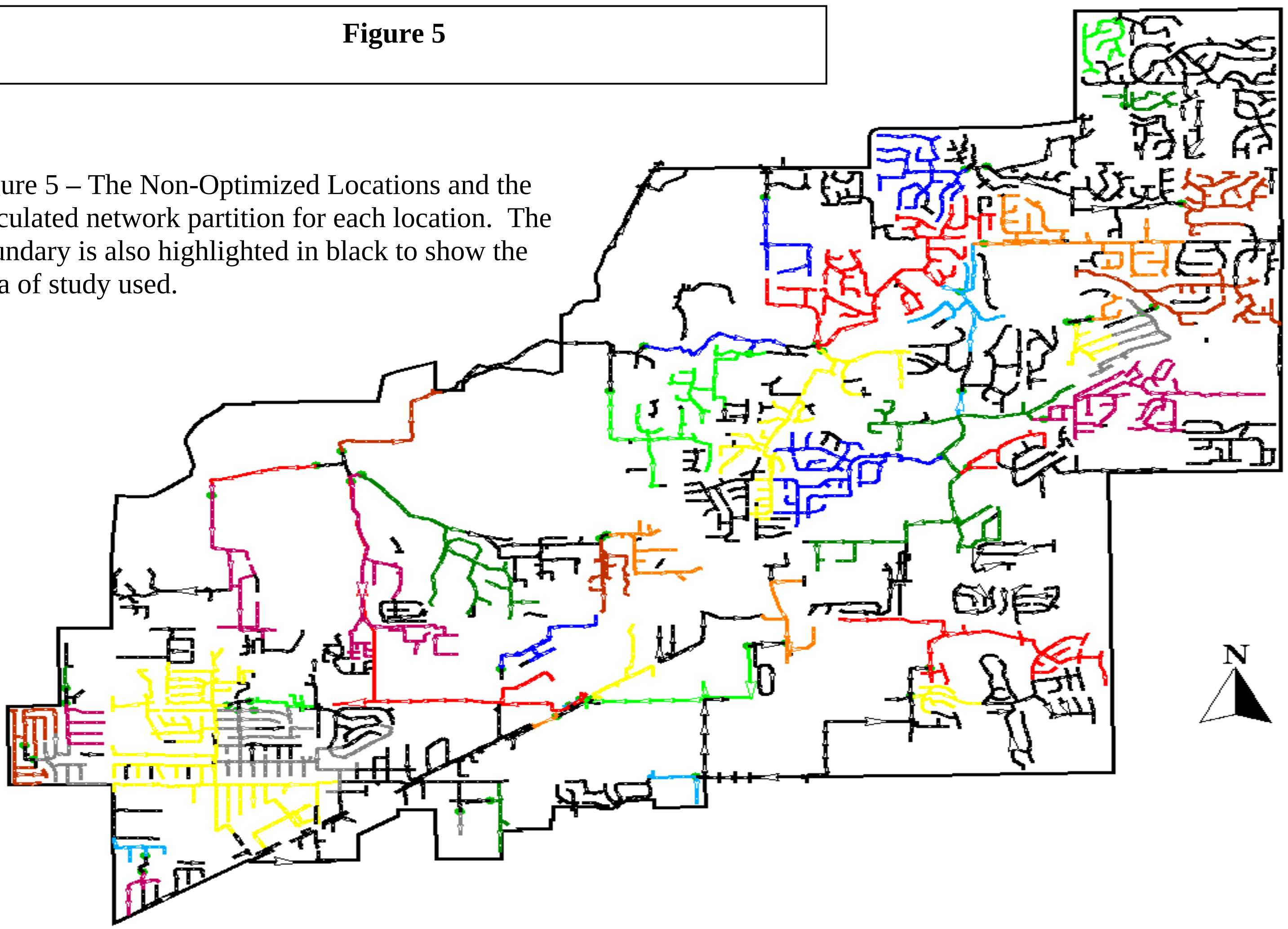


Figure 6

Figure 6 – The Optimized Locations and the calculated Network Partition for each location.

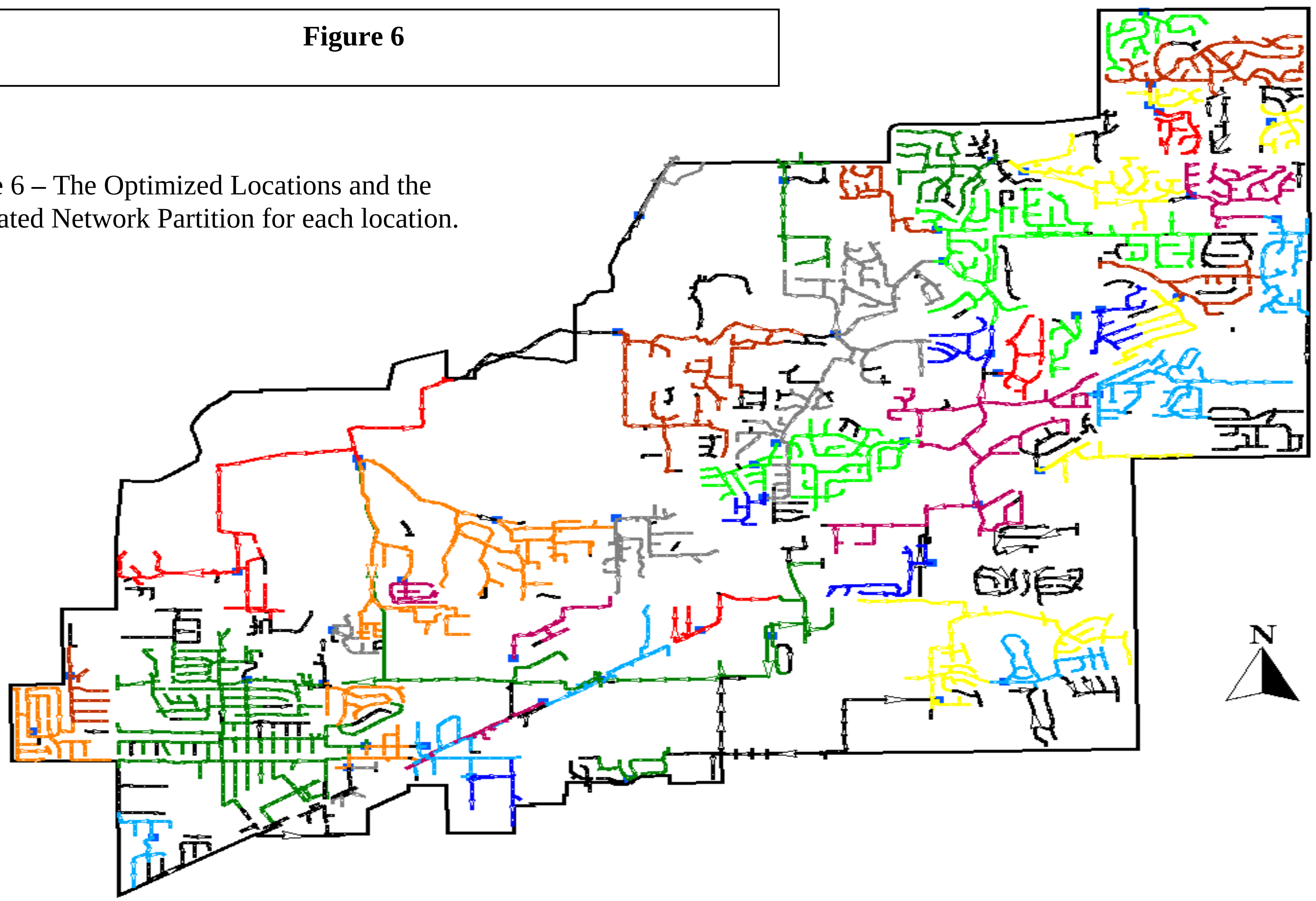
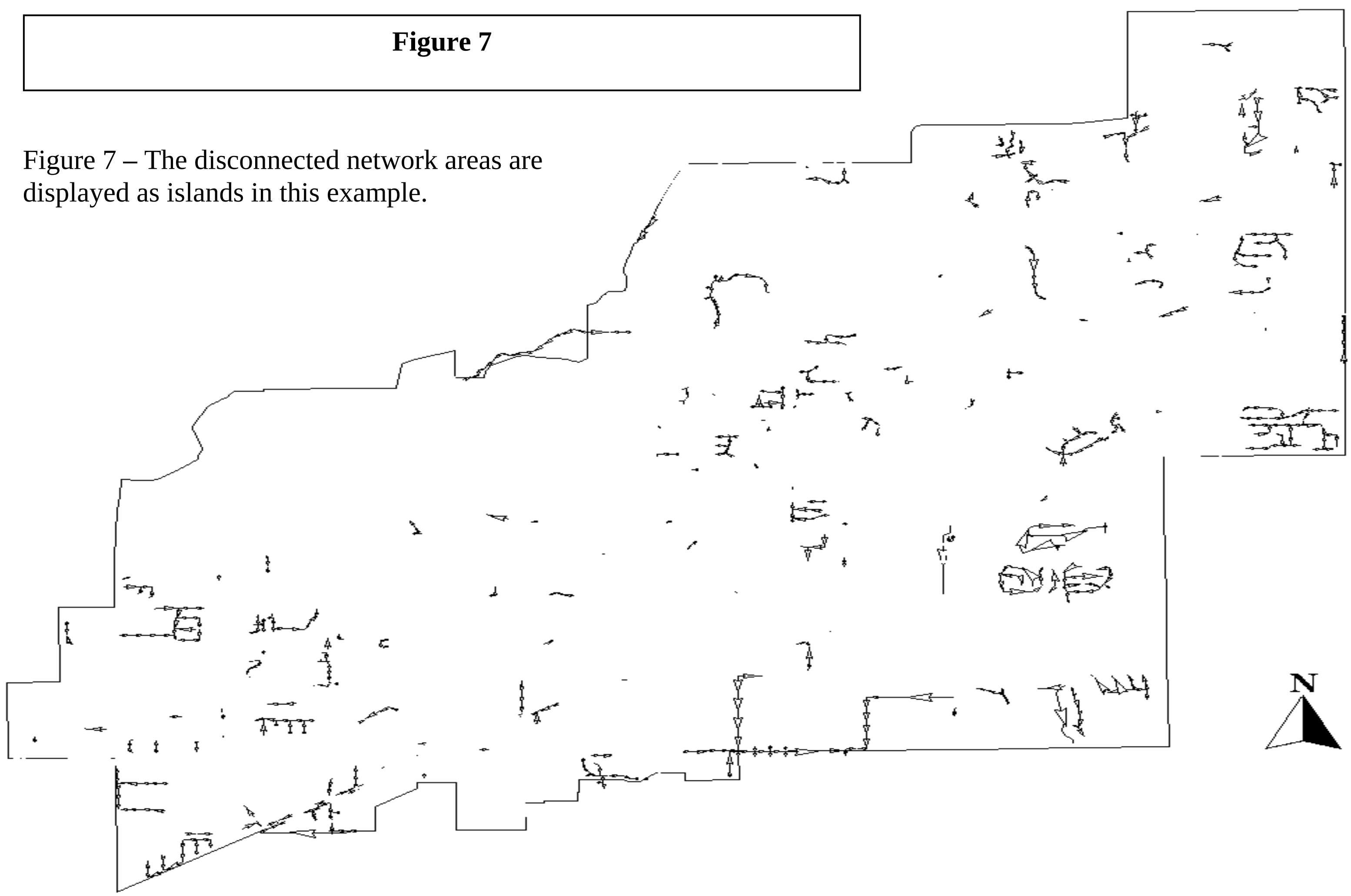


Figure 7

Figure 7 – The disconnected network areas are displayed as islands in this example.



References

Beven, K.J., *Rainfall-Runoff Modeling – The Primer*. John Wiley & Sons Ltd, 2000.

Biggs, N., Lloyd, E., and Wilson, R., *Graph Theory*, 1736-1936. Oxford University Press, 1986.

Christofides, N., *Graph theory: an algorithmic approach*. London ; New York ; San Francisco : Academic Press, 1975.

Churchman, C.W., Ackoff, R.L., and Arnoff, E.L., *Introduction to Operations Research*. J. Wiley and Sons, 1957.

Daskin, M.S., *What You Should Know About Location Modeling.* Wiley InterScience, 2008.

Densham, P.J., and Rushton, G., *A More Efficient Heuristic for Solving Large P-Median Problems*. Papers In Regional Science: National Center for Geographic Information and Analysis, Vol. 71, 1992.

ESRI Help Online, *ArcGIS Help Online*. Search Term: “Union Tool”. ESRI Online, 2009.

Gold, C.M., *Point and area interpolation and the digital terrain model, Trends and concerns of Spatial Sciences*. Second Annual International Seminar Proceedings, 1988.

Hajiaghayi, M.T., Mahdian, M., and Mirrokni, V.S., *The Facility Location Problem with General Cost Functions*. Wiley InterScience, Vol. 42, 2003.

ISDP Online, *Indiana Spatial Data Portal – Local Dataset*: 2003 Marion County LiDAR dataset. http://www.indiana.edu/~gisdata/local/marion2003_lidar.html (accessed November, 5 2009).

Kirby, M.W., *Operational Research in War and Peace: The British Experience from the 1930s to 1970*. Imperial College Press, 2003.

Larson, C., and Odoni, A.R., *Urban Operations Research.* Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

Massachusetts Water Resources Authority, *A History of the Sewer System, Online Website and Facts Sheet*. http://www.mwra.com/03sewer/html/sewhist.htm (accessed November 5, 2009).

Mirchandani, P.B., and Francis, R.L., *Discrete Location Theory*. John Wiley & Sons, Inc., New-York, 1990.

Mitchell, P.S., and Stevens, P.L., *Collection Systems 2005 – Sustaining Aging Infrastructure: System, Workforce and Funding*. Water Environment Federation, 2005.

Sanitary Sewer Overflow Analysis and Planning (SSOAP) Toolbox, *Urban Watershed Management Research*. http://www.epa.gov/ednnrmrl/models/ssoap/ (accessed November 5, 2009).

Stevens, P.L., *Basin Size is the Most Effective Variable an I/I Engineer can Control*. Excerpts from the 1993 WEF annual conference, 1993. (Updated in 2000 with additional data and case studies).

United States Environmental Protection Agency (US EPA), *Office of Water, Polluted Runoff (Nonpoint Source Pollution)*. http://www.epa.gov/nps/facts/point7.htm (accessed November 5, 2009).

USEPA, *Effects and Control of CSOs and SSOs*. Report to US Congress, 2005.

Zhang, Z., *Estimating Rain Derived Inflow and Infiltration for Rainfalls of Varying Characteristics*. Journal of Hydraulic Engineering, Volume 133, No. 1, January 2007.

Curriculum Vitae

Christopher A. Skehan

Education

MS, Geographic Information Sciences, 2010 – Indiana University, Indianapolis, IN
BS, Public Affairs, 2005 – Indiana University, Bloomington, IN

Professional Experience

A multitalented, organized, and analytical individual with a passion for environmental sciences and geographical information analysis. Ability to handle large scale project management both in the public sector and in the private sector. Effective group leader and experienced reporting writing expert. Superior skills and advanced knowledge of computer hardware and software, with expert abilities in GIS software and applications.

Major Achievements

At ADS Environmental Services, I have successfully worked my way into obtaining the GIS chair position, with emphasis on purchasing, training, and advising at a company-wide scale. I have managed data analysis projects that are over $1 million dollars in most of the major cities in the Midwest region of the United States. I have network and created partnerships between many different organizations to assist with the overall improvement of wastewater management and sewer collection system data in Indianapolis, IN. I provide annual training to our clients on most internal data analysis software developed at ADS Environmental Services.

At the Indianapolis Department of Natural Resources, I maintained, tracked, and updated the divisions PC inventory and assets and provided software and hardware support for all PC and network equipment for the division of State Parks. I also assisted in the creation of the DNR helpdesk, which now serves over 300 DNR state employees.

Honors and Awards

As a graduate student at IUPUI, during the course of my MS in GIS, I maintained a overall academic GPA of 4.0 out of 4.0. I have been awarded the Indiana University Deans List award in 2004, 2005, 2008, 2009, and 2010. During my undergraduate career, I was awarded the Rabes Vermillion Scholarship in 2000 and 2001.

Community Involvement

Served on the Franklin Lake Board of directors as the Sectary from 2008 through 2010. This is a neighborhood association that resides in the Franklin Township of Marion County Indiana.

Certifications

Graduate Certification in GIS 2009

PACP and MACP Certification in 2009